

Altering the Oligomerization State of GFP-Like Purple Protein to Enhance Protein Tagging Ability

Samantha Kuhn

University of Texas at Austin
Dean's Scholars Honors Thesis
2009

Scott Stevens, Ph.D.
Molecular Genetics and Microbiology
Supervising Professor

Date

Table of Contents

Abstract	Pg. 3
Background	Pg. 5
Methods and Materials	Pg. 15
Results	Pg. 22
Discussion	Pg. 29
Appendix	Pg. 31
References	Pg. 33
Acknowledgments	Pg. 35

Abstract

Since its discovery and purification in the 1960s, Green Fluorescent Protein has quickly become an extremely useful scientific tool for analyzing protein expression and dynamics. GFP was originally discovered in a species of jelly fish, *Aequorea victoria*, apparent only in response to a blue flash of bioluminescence. Upon examining its structure, scientists soon found that the source of GFP's color was an autocatalytically formed chromophore at the center of its β -barrel structure. Because of this autocatalytic mechanism of fluorescence, the GFP gene alone can be attached to proteins of interest and used to track expression and movement in vivo.

While GFP itself was a breakthrough in the biotechnology world, it is hardly the end of the line as far as fluorescent proteins go. In recent years, scientists have been searching for and finding more members of the GFP-family that have the same autocatalytic property but display different colors. The introduction of new colors to the library of fluorescent proteins offers great advantages in multiple protein tagging and studying protein interactions. However, most fluorescent proteins that have been found to date have been tetramers, consisting of four identical subunits. This quaternary structure makes them ineffective as protein tags.

Researchers have recently been exploring ways to alter the oligomerization state of fluorescent proteins and create dimeric and monomeric forms that can be useable tags. In my experiment, I attempted to do so in a GFP family protein from species *Galaxea fascicularis* and *Montipora efflorescens* known as purple protein. There is not a protein tag that emits is colored purple currently, and so my goal was to reduce the oligomerization state of purple protein, assuming that it is a tetramer, while maintaining its purple color. To do this, I used site-directed mutagenesis techniques to change specific amino acids in the protein that were predicted to be involved in formation of tetramers.

All three mutations made in the sequence affected the spectral qualities, as the cells they that the mutants were transformed into showed no purple expression. One mutation resulted in a protein that migrated at half the relative molecular weight of the wild type protein in a gel filtration column, and thus we were successful in reducing the oligomerization state of the purple protein through that mutation. While it is not currently useful as tag, the successful mutant should be used in future experiments to isolate a functional monomer of the purple protein, adding to the library of protein tag colors.

Background

Introduction

Green Fluorescent Protein (GFP) was first discovered in a species of jelly fish, *Aequorea victoria*, and purified in the 1960s. Scientists observed that, when exposed to blue light, the jelly fish emitted a green fluorescence. (Matz, Lukyanov, & Lukyanov, 2002) After isolating the protein, intensive study of its structure and function led to revolutionary discoveries that translated into breakthroughs in biotechnology, particularly with respect to the observation of protein expression and interactions. While GFP initially seemed to be one of a kind, researchers have recently been in pursuit of an entire library of GFP-like proteins – those that display similar properties via the same mechanism as GFP, but differ in structural and spectral properties. (Matz, Lukyanov, & Lukyanov, 2002) These GFP-like proteins offer the possibility of greater innovation in protein tracking techniques *in vivo*. However, a typical problem that has come up in the study of these proteins is oligomerization (Shaner, Patterson, & Davidson, 2007). Due to the tetrameric quaternary structure of the majority of these proteins, their natural state is not conducive to using them as fluorescent tags for protein interaction and expression. (Tasdemir, et al., 2008) Therefore, there is a great deal of research to be done on manipulating the structural character of many of these proteins to make them better tools for biologists and biochemists.

Green Fluorescent Protein

Green Fluorescent Protein (GFP) is one of many fluorescent tools that have been used for tracking protein expression and interaction; however it has been recognized as one of the most valuable. In the past, scientists have employed small organic fluorophore dyes for tracking proteins, but due to the dyes' lack of specificity, antibodies are required to target proteins of

interest. This requires the cells to be fixed and permeabilized, thus preventing the tracking of proteins in living cells. Following that, phycobiliproteins were the first fluorescent proteins used in cell biology. Phycobiliproteins are photosynthetic pigments found in cyanobacteria. These macromolecules are brighter than the dyes since each protein contains multiple chromophores, but their use is limited by their large size. Around 200 kDa, phycobiliproteins do not diffuse very well and have thus been mainly used for surface protein labeling. (Giepmans, Adams, Ellisman, & Tsien, 2006)

With the discovery of GFP, many of the obstacles of protein expression imaging were overcome. The new benefits of GFP were predominantly the result of its unique structure and mechanism for generating fluorescence. The basic structure of a single GFP monomer is an 11-stranded β -barrel with an α -helix running through the center of the cylinder [Figure 1-A]. (Yang, Moss, & Phillips, 1996) Unique to GFP (and the GFP-like family of proteins), the formation of the chromophore occurs autocatalytically during a post-translational protein maturation stage. (Pakhomov & Martynov, 2008) Requiring only oxygen, inward-facing β -barrel sidechains containing a Serine-Tyrosine-Glycine sequence spontaneously react to create the basic GFP chromophore molecule: 4-(*p*-hydroxybenzylidene)-5-imidazolinone, referred to as *p*-HBI [Figure 1-B]. (Giepmans, Adams, Ellisman, & Tsien, 2006)

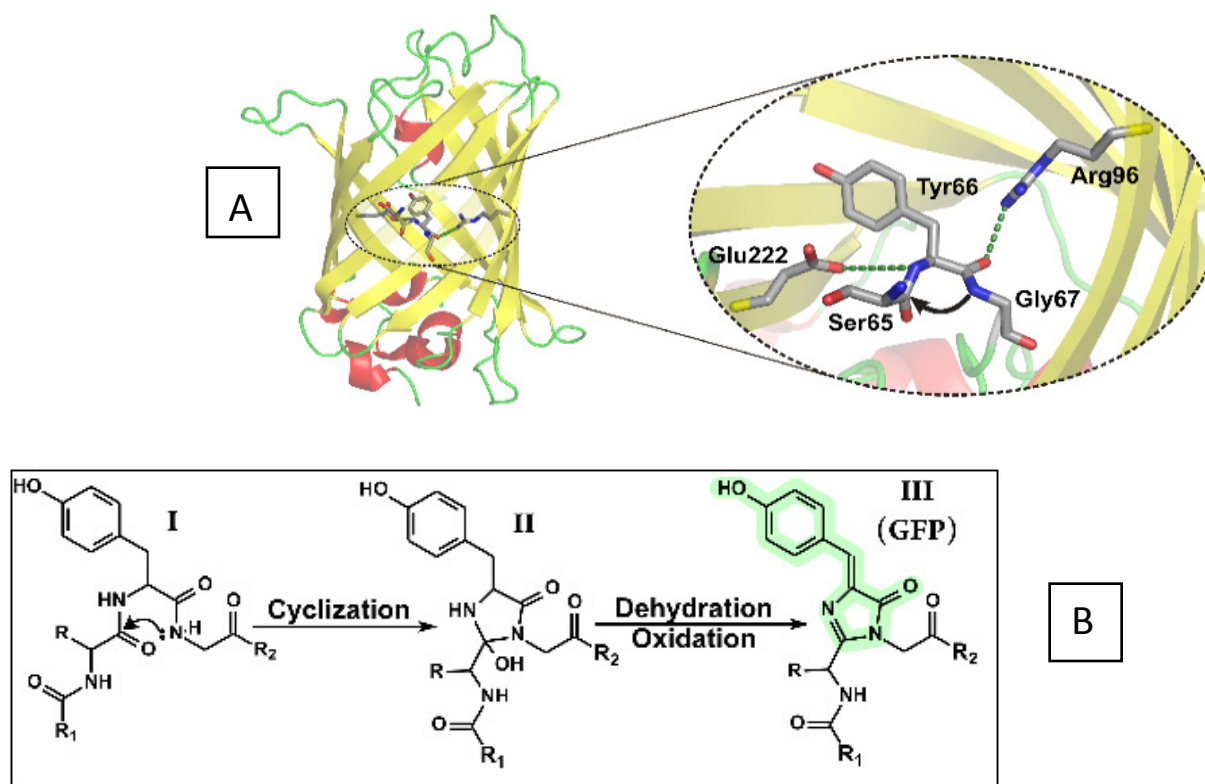


Figure 1. (A) Green Fluorescent Protein Structure: β -barrel, α -helix, and internal amino acid residues; (B) Cyclization and Oxidation reactions to create GFP basic structure, *p*-HBI, from tripeptide sequence [Borrowed from *GFP Family: Structural Insights into Spectral Tuning*]

In addition to being able to serve as a reporter gene or selection marker (Tasdemir, et al., 2008), the structure and functional properties of GFP make it an ideal tool for genetic protein tagging. It can be covalently fused to just about any protein due to its bright fluorescence, autocatalytic chromophore formation, and the fact that it is genetically encoded. This allows for precision in tagging, easy delivery to cells or organisms, and the ability to observe protein localization and expression *in vivo*. (Giepmans, Adams, Ellisman, & Tsien, 2006) In this respect, the uniqueness of GFP makes it an important subject of study in the field of biotechnology.

GFP-like Proteins

As with any great discovery, the finding of GFP led to an explosion of research and attempts to improve/expand the applications of GFP. Initially, researchers worked to genetically alter GFP itself to produce other colors, as well as brighter, more versatile versions of itself (e.g. reduced oligomerization). They were successful in creating blue, cyan, and yellow variants of GFP. (Shaner, Patterson, & Davidson, 2007) The second aspect of the pursuit, however, was the search for other fluorescent proteins (FPs) to be isolated from other organisms but having similar fluorescence as GFP. Ideally, scientists wanted to find FPs with spectral properties that made them distinct (particularly, ones that emitted wavelengths that were distinguishable from the current GFP-like proteins available). For a long time, scientists focused on bioluminescent systems only, expecting their similar physical properties to be indicative of similar structural bases. However, many of these organisms were evolutionarily distant from the jelly fish species containing GFP, and in fact used systems far removed from that of GFP. It was not until they branched out and explored non-bioluminescent organisms, such as corals, that they found natural GFP-like proteins that possessed the fluorescent capabilities and were visibly colored. (Matz, Lukyanov, & Lukyanov, 2002) This breakthrough was followed by the discovery of several other FPs, most of which were non-bioluminescent but also fluorescent and colored, responsible for the pigment in the organisms that expressed them (Lukyanov, et al., 2000).

Members of the GFP family of proteins are similar to GFP, as would be expected, in their basic structure and mechanism of fluorescence. They contain the same β -barrel structure surrounding the chromophore-containing α -helix. The same autocatalytic mechanism is used to produce the chromophore, resulting in their color (linking them to GFP). (Matz, et al., 1999) What makes them appear different colors, however, is the chromophore structure itself. It

appears that during chromophore synthesis, different organisms produce alternative chromophore products that have varying spectral qualities. (Labas, et al., 2002) It appears that the diversity results from chemical modifications to the *p*-HBI core after it has been synthesized. (Pakhomov & Martynov, 2008)

Importance of New Fluorescent Proteins

The technological implications of discovering these new, non-green FPs are very great. To reiterate, using GFP for in vivo labeling has quickly become one of the most popular biochemical techniques for analyzing protein expression and behavior in cells. One of the big benefits of using genetic engineering to label proteins of interest within cells is that the label can have very little to no effect on the function of the protein or the cell itself. (Schultz & Kohn, 2008) Until the discovery of GFP-like fluorescent proteins, however, technological innovation was very restricted because the library of colors available was limited to GFP and the mutant variants (e.g., cyan, yellow, and blue). While the mutants allowed for some advancement, such as dual-color labeling and fluorescence resonance energy transfer (FRET) between colors for studying protein-protein interaction, there was not much beyond this that could be accomplished. (Matz, Lukyanov, & Lukyanov, 2002)

The expansion of fluorescent protein use due to the discovery of GFP-like proteins applies to single protein analysis as well as protein-protein interactions within cells. I will briefly examine the techniques that have developed and/or benefited from the discovery of non-green FPs before explaining the importance of adding to the library of usable fluorescent proteins. The first aspect is the ability to analyze conformational changes within an individual protein. Using fluorescence resonance energy transfer (FRET), scientists can flank a protein

between two marker proteins (i.e. two FPs of different spectral qualities) to detect changes in the conformation of the protein, such as if it undergoes phosphorylation. The change is indicated by a shift in the FRET between the two FPs. (Giepmans, Adams, Ellisman, & Tsien, 2006)

The GFP family of proteins offers several advantages to protein-protein interaction observation. For one, bimolecular fluorescent complementation (BiFC) is a technique that is used to study whether certain proteins interact with one another, and when that occurs. The technique is possible because of the ability of GFP family proteins to reform their functional chromophores after being split and reunited. In BiFC, an FP is split and each half is tagged to a different protein (typically two proteins that are thought to interact). When the proteins are expressed, if/when they interact, the FP will come together and fluoresce, indicating the location and time of interaction. Additionally, tagging multiple proteins with different colors allows us to examine protein interaction when they act in conjunction with one another (multi-color fluorescent correlation spectroscopy), for example, if two proteins diffuse simultaneously or in the same region. (Giepmans, Adams, Ellisman, & Tsien, 2006)

While the mutant variants of GFP have allowed for these innovations, finding additional colors will allow for even greater advancements. For one, having a variety of FP reporters of different wavelengths will allow for versatility when it comes to looking at protein expression in different types of tissue and cells. New colors offer potentially better visibility in cells that have problems with the expression of these proteins. Additionally, multi-coloring makes it possible to observe multiple proteins at once being expressed or moving in the cells, thus saving time and money in the lab. (Matz, Lukyanov, & Lukyanov, 2002) This also applied to BiFC experiments, where multiple interactions can be observed at once. Finally, new fluorescent proteins provide researchers with more FRET partner molecules, thus allowing for the study of higher order

complexes (i.e. protein interactions and the conformational changes that occur during the interactions). (Giepmans, Adams, Ellisman, & Tsien, 2006)

Engineering of GFP-like proteins

While the GFP family of proteins offers great advantages, there is one major problem that has been encountered. So far, all proteins that have been investigated form very tight oligomers (usually consisting of at least 4 monomers, i.e. tetramers). The complex quaternary structures of these proteins limits their use in FRET techniques (Matz, Lukyanov, & Lukyanov, 2002) and significantly limits their use as protein fusion tags. (BD Biosciences, 2005) Therefore, a major task for researchers involved in this field has been to find ways to prevent the oligomerization that occurs between the protein monomers (or at least reduce it) while retaining the fluorescent color expression. This has been primarily attempted through modification of the amino acid sequences of the proteins through mutagenesis of the wild type DNA. (Matz, Lukyanov, & Lukyanov, 2002)

In recent years, there have been successful attempts at genetically altering fluorescent proteins to reduce their oligomerization states. One important achievement was the alteration of protein drFP583, also known as DsRed. An interesting fact about DsRed is that it was the first known fluorescent protein that changes its emission color over time, making it an excellent indicator of time elapsed since a protein began being expressed when used as a reporter gene. (Matz, Lukyanov, & Lukyanov, 2002) The problem in its utility was that it, as usual, is a tetramer in nature. Its strong oligomerization prevents it from being a useful protein tag (Baird, Zacharias, & Tsien, 2000) However, scientists have been able to isolate the protein and mutate it to form a functional Ds-Red monomer, which is now marketed as a red fusion tag. (BD

Biosciences, 2005) A similar protein, HcRed, has also been converted from tetrameric form to a dimer. (Gurskaya, et al., 2001) The methodology behind creating these monomers involves making site-specific amino acid mutations in the wild type proteins in order to disrupt those involved in oligomerization and prevent tetramerization from occurring. This often requires several mutation attempts, and has required additional mutations to recover lost functions (e.g. if mutations prevent oligomerization but also eliminate or alter the spectral properties of the protein). (Tasdemir, et al., 2008)

Purple Protein

The protein I have been focused on is a non-bioluminescent and non-fluorescent purple protein. The two versions of the protein we have been working with come from the species *Galaxea fascicularis* and *Montipora efflorescens*. This protein was discovered by Mikhail Matz, and it is assumed to be a tetramer due to the typical nature of these types of proteins. When expressed, the purple protein gives cells a bright purple color.

I decided to work with the purple protein because a purple colored protein has not yet been isolated and made viable for protein tagging, so my goal in this project was to introduce mutations into the purple protein sequence that would alter the oligomerization state. In one respect, we hoped to find out/confirm if the oligomerization state of natural purple protein is actually a tetramer. Additionally, assuming that it was a tetramer, we were attempting to make mutations that would prevent tetramerization and render it either a dimer or monomer while still maintaining the purple fluorescence of the wild type protein. If a monomer (or dimer) was successfully engineered and exhibited the purple color in expression, we hoped that it could be made into a useful protein tagging tool to be added to the current library of monomeric

fluorescent proteins. Additionally, it would be a useful tool for following the purification of tagged recombinant proteins from *E. coli* or other organisms.

Project Overview

The project was designed so that three specific mutations would be made in the wild type purple protein, and that combinations of these mutations would also be made, so that there would be seven total mutant genes: three single mutants, three doubles, and one triple. Deciding on the mutations to make could not be done solely based on the purple protein itself because the exact structure is currently unknown. Therefore, the mutations were chosen by comparing the sequence of purple protein to that of a homologous protein, and using the structure of the homolog as the guide. [Figure 2]

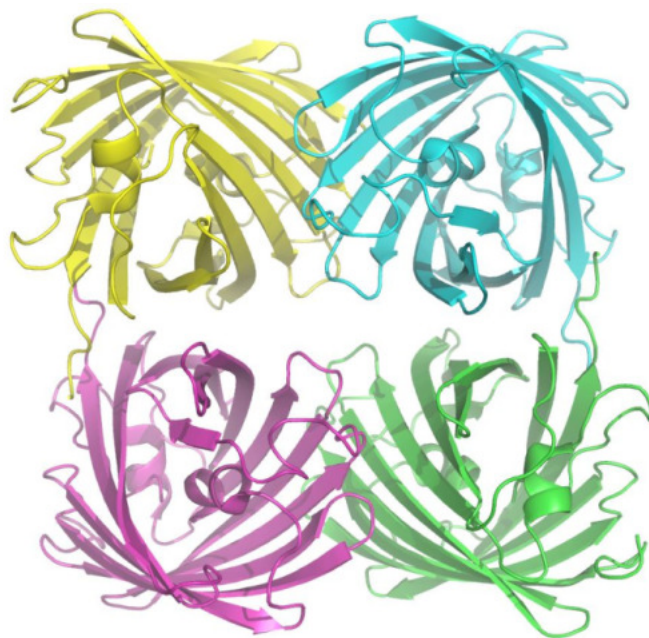


Figure 2. Structure of a tetrameric prote in homolog of purple protein. (PDB ID: 2P4M)

The three residues selected to be mutated were amino acids known to be involved in the tetramer formation in the purple homolog (e.g., the formation of salt bridges) [Figure 3]:

- 1) Threonine at amino acid position 102 was changed to a lysine (mutation T102K).
- 2) Arginine at amino acid position 149 was changed to a glutamic acid (mutation R149E).
- 3) Phenylalanine at amino acid position 158 was changed to a lysine (mutation F158K).

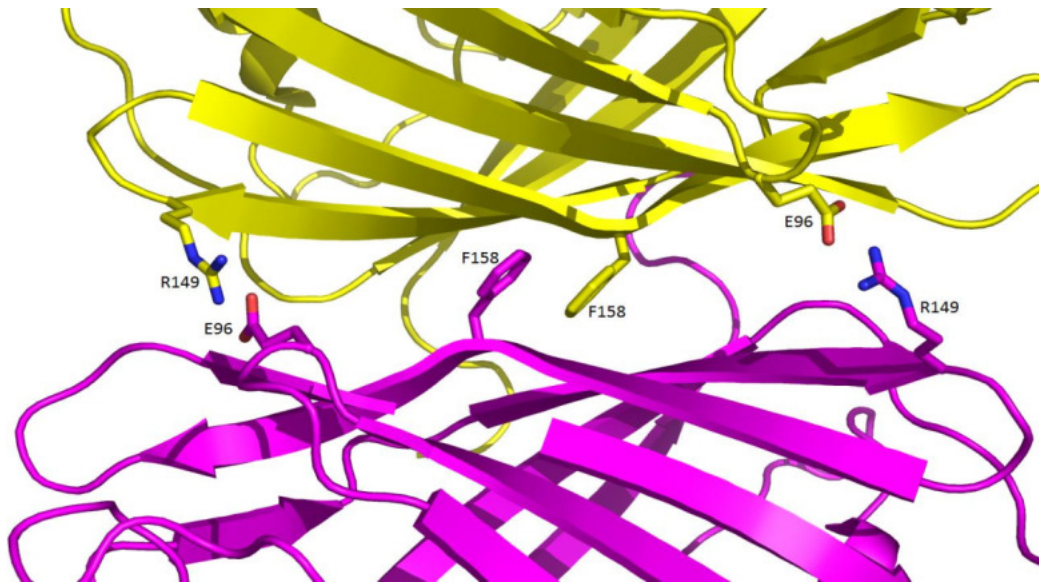


Figure 3. Image of tetrameric interactions of purple protein homolog. The involvement of residues F158 and R149 can be seen.

The plasmid chosen as the purple protein vector was pET-16b, which contains an ampicillin resistance gene for screening purposes. The vector also has a HIS-tag so that it can be purified and analyzed after mutagenesis and transformation. Once the mutations were made, the plan was to transform the mutant-containing plasmids into high-efficiency protein expression cells, and then check for expression of the purple color and isolate the mutant proteins for relative molecular weight analysis.

Methods and Materials

Site-Directed Mutagenesis

Mutations were introduced into the purple protein gene through PCR reactions using site-specific mutant primers. Two different vectors containing the gene were used as templates for the PCR reactions: pET-16b and pET-22b(+). Primers for the three mutations were engineered to be identical to the purple protein gene with the exception of the single codon that coded for the amino acid to be changed (see Appendix for primer sequences). In creating multiple mutations on one gene, single mutants were used as the templates for double mutants, and a double mutant was used as the template to introduce a third mutation.

The PCR reactions consisted of 10X PCR reaction buffer, 10X dNTPs, 125 ng of forward primer, 125 ng of reverse primer, and 100 ng of template. One unit of Pfu DNA polymerase was used per reaction, and each reaction was a total of 50 μ l in a PCR tube. The times and temperatures for the program used were the following:

1. 95° C for one minute
2. 18 cycles of: 50 second denaturing at 95° C, 50 second annealing at 57.6° C, 9.5 minute polymerization at 70° C
3. 70° C for seven minutes
4. Hold at 4° C until removed

Note: 57.6° C was used as the annealing temperature for most mutations. However, when this temperature failed to produce product, a gradient of annealing temperatures was used, consisting of one at each of 68° C, 63.3° C, 57.6° C, and 55° C.

DpnI Digest

In order to eliminate the template before transforming the products into *E. coli*, each PCR reaction was digested with DpnI. For each 50 µl reaction, 0.5 µl (10 units) of DpnI was added. The contents were mixed using a pipette, and then each tube was centrifuged for one minute at 4000 RPM to pool the contents at the bottom. The mixtures were then incubated in a 37° C water bath for an hour.

Gel Electrophoresis

To analyze DpnI digestions (as well as restriction digests), analytical agarose gels were used, composed of 1% agarose and 50 ml of 1X TAE buffer with 1 µl of 10 mg/ml ethidium bromide. Typically, 10 µl of each sample was combined with 2 µl of 5X loading dye. Each gel was run at a constant voltage of 100V for one hour, with the samples alongside a 1 KB size standard.

XL-10 Gold Transformation

Each PCR product was transformed into XL-10 Gold *E. coli* cells to screen for mutants. 100 µl of XL-10 Gold cells was added to 15 µl of PCR product, which was then incubated on ice for 30 minutes. All 115 µl were plated on 2XYT + Ampicillin plates and incubated at 37° C for 16-24 hours.

DNA Purification

For each transformation, isolated colonies were picked and inoculated in 3 ml of 2XYT medium containing 450 µg of Ampicillin. These cultures were grown 16-24 hours. The next

day, 1.5 ml of each culture was centrifuged at 5000 RPM for one minute. The DNA was purified from the cultures using Sigma DNA Miniprep Kit.

Restriction Digest

Restriction digests using enzyme PvuII were done to assay for the expected DNA fragment sizes (based on the non-mutant purple gene contained in the plasmids as a control). The reactions contained 10X NEB Buffer 2, three units of enzyme PuvII, and 3 µl of the miniprep DNA. Each digest contained a total volume of 10 µl. (The volume was brought up to 10 µl with double-distilled water.) The reactions were then incubated in a 37° C water bath for one hour. After incubation, 2 µl of 5X loading dye was added to each sample, and all 12 µl were run on a 1% agarose gel.

DNA Sequencing

Using the restriction digest for comparison to the control, clones that looked to be positive for the plasmid vector were sequenced. 500 ng of each DNA miniprep was aliquoted and sequenced by the ICMB sequencing facility using T7 primers. The sequences were compared to the expected mutant sequences for verification of mutations.

Protein Expression

Verified mutant DNA strains were transformed into BL21 (DE3) *E. coli* cells for high-efficiency protein expression. 1 ng of each mutated construct was added to 25µl of BL21 (DE3) cells, and the transformation reaction was incubated on ice for 30 minutes. The mixtures were then plated on 2XYT + Ampicillin plates and incubated at 37° C for 16-24 hours. The following

day, an isolated colony from each plate was streaked onto 2XYT + Ampicillin + IPTG plates to induce expression. (Prior to colony streaking, 8 μ l of 1 M IPTG was spread on the surface of the plates.) The plates containing IPTG (as well as non-IPTG streak plates, used as controls) were incubated at 37° C with exposure to white light for 16-24 hours.

IPTG Concentration Gradient

To test for the best concentration of IPTG in inducing purple protein expression, a time course was set up using the non-mutant purple control and the triple mutant as test samples. Five overnight cultures were grown (containing 3 ml 2X YT + 450 μ g Ampicillin) for both the control and the mutant clones. The following day, each culture's absorbance was measured and diluted to an OD of 0.10 at 600 nm (A_{600}) in new media. They were allowed to grow to an absorbance of about 0.45, at which point they were induced with one of five concentrations of IPTG: 0.05 mM, 0.10 mM, 0.25 mM, 0.50 mM, and 1.00 mM. The new cultures were incubated in light at 37° C for 16-24 hours.

Protein Purification

To test for the oligomerization state of the wild type and mutant proteins, the expressed proteins were purified from the cells. A colony from each of the plasmid-containing BL21 (DE3) cell transformations were inoculated in 3 ml of 2XYT + 450 μ g Ampicillin media (mutants inoculated include all three single mutants and the triple mutant, as well as the wild type). After growing the cultures for 18 hours, 2 ml of each culture was added to 200 ml of 2XYT media + 600 μ g Ampicillin in 1-liter flasks. The cultures were grown to an OD of about

0.45, at which point IPTG was added at a concentration of 0.1 mM (20 µl of 1M IPTG). These cultures were grown overnight (about 18 hours).

The next day, each culture was centrifuged at 5000 RPM for 10 minutes in Oak Ridge tubes. The supernatant was removed from the pellet, and the pellet was resuspended in a buffer (20 mM NaPO₄ and 200 mM NaCl; pH 7), which was then transferred to a 50 ml conical vials and flash-frozen by submerging in liquid nitrogen. In order to lyse the cells, the pellets were thawed in ice water and refrozen in liquid nitrogen a total of three times. After the final thaw, the cells were sonicated using the full duty cycle on setting 7 for a total of 5 cycles (15 second sonication followed by 45 seconds of rest for each time). Finally, the cells were refrozen once again in liquid nitrogen and stored in the -80° C freezer.

The frozen cells were thawed in ice water. 2 µl of the whole cell suspension was aliquoted into a 1.5 ml Eppendorf tube (to run on a gel later) and preserved in the -80° C freezer. 1.5 ml of 10% Triton X-100 was added to each batch of cells. The lysed cells were centrifuged at 15,000 RPM for 20 minutes. 2 µl of the supernatant was removed and saved in the -80° C freezer.

To purify the protein, Biorad Poly-Prep Chromatography Ni-NTA columns were used. All procedures were performed in a 4° C room, and 5-10 µl aliquots of each solution that flowed through the column (the sample, wash buffer, and eluted fractions) were removed and saved in the -80° C freezer to run on a gel. First, the columns were prepared by adding 400 µl of Qiagen NiNTA beads, followed by 10 ml of a Column Preparation Solution (Resuspension Buffer, 10 mM β-mercaptoethanol). The supernatants from centrifugation were then passed through the column. Then, each column was washed three times with 10 ml Wash Buffer (Resuspension Buffer, 10 mM β-mercaptoethanol, 25 mM Imidazole; pH 8). The protein was eluted five times

with 200 µl of an Elution Buffer (Resuspension Buffer, 300 mM imidazole) into five separate 1.5 ml Eppendorf tubes. The 200 µl fractions were stored in the 4° refrigerator.

Protein Gel

Samples taken from the flow through of the column were run on a Novex 4-20% and NuPAGE gel. 20 µl of 2X LDS Buffer was added to 2 µl of each of the whole cell suspension, soluble protein sample, and the sample flow through from the column. 5 µl of the buffer was added to 5 µl of each eluted fraction (in actuality, only the two most intensely purple – i.e. where the purple protein was most concentrated – fractions were run on the gel). Therefore, for each mutant protein (and the wild type), five total samples were run on the protein gel (2 fractions, the whole cell, the soluble protein, and the flow through). The samples were run alongside Invitrogen Benchmark Protein Ladder. The gel was electrophoresed in SDS-PAGE protein running buffer at 200V for 41 minutes.

After electrophoresis, it was removed from its casing and fixed by placing it in a small container and submerging it in a solution of 50% methanol and 10% acetic acid. The gel was kept in this solution for 15 minutes on a shaking plate. Then, the gel was washed for 10 minutes (shaking) in ddH₂O. After washing, the gel was stained with a solution containing 20% methanol, 2% phosphoric acid, 10% ammonium sulfate, and 0.05% Coomassie G-250 overnight.

Sephadex S200 Column Filtration

In order to determine the relative molecular weights of the mutant proteins (i.e. the oligomerization state relative to the wild type purple protein), a Sephacryl S200 Column was

used to separate the different sizes via gel-filtration. The column separates proteins less than 200 kDa. One of the fractions for each mutant was selected and loaded onto the column. Using Resuspension buffer, the fraction was passed through the column, and 2.5 ml fractions were collected. A chart recorded the level of protein that was eluted into each fraction, indicating the relative sizes of the proteins in each sample.

Western Dot Blot

A second test for protein size that was performed was the western dot blot. From the Sephacryl S200 Column fractions, 250 µl from every third fraction for each mutant was aliquoted into a 1.5 ml Eppendorf tube. A dot microfiltration apparatus was used to perform the dot blot, so in preparation, a sheet of nitrocellulose paper was cut to the size of the apparatus. The paper was washed for five minutes in ddH₂O, and then washed for another five minutes in TBS (20 mM Tris –pH 7, 150 mM NaCl). It was then set on top of the membrane, and the 96 well plate was placed on top (and tightened). The manifold was hooked up to a vacuum. Each sample was loaded into one of the 96 wells, including a sample of the pure purple protein fraction (that had not been passed through the Sephacryl column) as a control. The primary and secondary antibodies were incubated using the protocol for chemiluminescent substrates in the QIAexpress Detection and Assay Handbook. The first antibody was diluted about 1:1300, and the second antibody used was goat anti-mouse IgG conjugated with horseradish peroxidase at a 1:5000 dilution. The signal was detected using chemiluminescence as described by the manufacturer (Perkin Elmer). After completing the transfer, the dot blot was exposed on X-ray film using three different exposure times: 10 seconds, 30 seconds, and 1 minute.

Results

Mutagenesis

At the beginning of the project, we were using the pET-16b plasmid as the vector for purple-protein. The first attempts at mutations were made using purple in pET-16b as the template [Figure 4]. After several successful mutations were made, we realized that purple was not being expressed in the transformed cells. Assuming that it was a problem with the vector, we switched over to the pET-22b(+) vector and began redoing the mutations. However, when we came across the same problem with the new construct, we were able to determine that the problem was with the cell line we were using. For that reason, while we were able to introduce the complete set of mutations (all singles, doubles, and triple), the library of mutants consists of some in pET-16b and others in pET-22b(+). In addition, the majority of PCR reactions were successful using an annealing temperature of 57.6° C, but we were unsuccessful creating the triple mutant at this temperature, and thus had to use a gradient. We ended up getting the best result at 68° C annealing temperature. Table 1 shows a complete list of the mutations made, the plasmids that they are in, and the annealing temperatures that worked for each.

Mutation	Vector	Annealing Temperature
T102K	pET-22b(+)	57.6° C
R149E	pET-16b	57.6° C
F158K	pET-22b(+)	57.6° C
T102K & R149E	pET-16b	57.6° C
T102K & F158K	pET-22b(+)	57.6° C
R149E & F158K	pET-16b	57.6° C
T102K & R149E& F158K	pET-16b	68.0° C

Table 1. Vectors and Annealing Temperatures for Gene Mutations

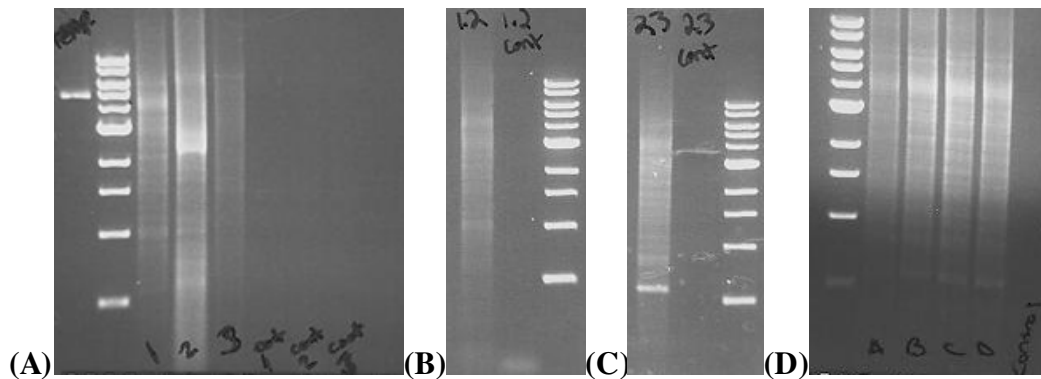


Figure 4. DpnI digests of PCR reactions for mutations in pET-16b plasmid. (A) The far left lane contains purified template of pET-16b containing the purple gene. To the right of the 1 KB ladder are the PCR reactions of mutations (from left to right) T102K, R149E, and F158K. The controls on the far right show PCR reactions done without DNA polymerase. (B) The PCR reaction and control for the double mutant: T102K & R149E. (C) The PCR reaction and control for the double mutant: R149E & F158K. (D) The PCR reaction gradient for the triple mutant, where A, B, C, and D correspond to annealing temperatures of 68.0°C, 63.3°C, 57.6°C, and 55.0°C, respectively (alongside the control in the far right lane).

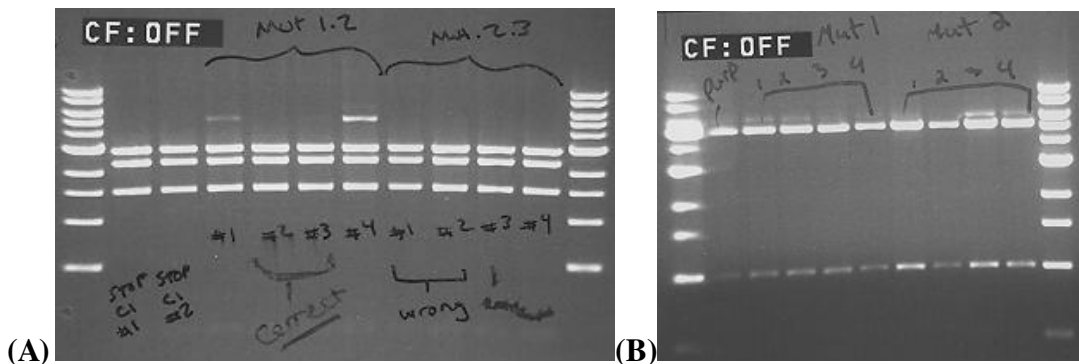


Figure 5. PvuII digests of plasmids after transforming constructs into XL-10 Gold *E. coli* cells and then purifying the DNA. Each gel shows a 1 KB ladder followed by the non-mutant plasmid containing purple for comparison. (A) Sample of digests of pET-16b clones. “Mut 1.2” designates double mutation T102K and R149E. “Mut 2.3” designates double mutation R149E and F158K. (B) Sample of digests of pET-22b(+) clones. “Mut 1” designates mutation T102K. “Mut 2” designates mutation R149E.

Using the PvuII restriction enzyme, we digested mutated constructs (purified from transformed XL-10 Gold *E. coli* cells) for comparison to non-mutants to make sure the correct plasmid was present. The gels of the digests [Figure 5] showed us if there were any unexpected

bands or sizes of DNA fragments that resulted from mutations. We chose certain clones for each mutant to sequence, and those that were correct were used for the high-efficiency protein expression transformations. (See Appendix for sequences of the wild type purple gene and the theoretical mutant sequences.)

We initially transformed our mutants and original plasmid into Rosetta Blue (DE3) *E. coli* cells. After the first attempt showed no expression of purple protein by the positive control, we switched vectors (as stated before) and redid the mutagenesis and transformations. When the problem persisted, we proceeded to do several tests. We grew the cells on 2XYT + Ampicillin + Chloramphenicol plates containing IPTG and no IPTG. The cells grew (without expressing purple) on the non-IPTG plates, but showed no growth on the plates containing IPTG. We then tried to grow the cells in liquid media, inducing with varying concentrations of IPTG at an absorbance of OD = about 0.3, but still could see no positive results for the control. At this point, we decided to switch to a different cell line – BL21 (DE3) *E. coli* cells.

Purple Protein Expression

Using the previously constructed mutants, we transformed all plasmids into BL21 (DE3) cells. The result was that, on 2XYT + Ampicillin plates with IPTG, the purple control expressed the purple protein, but none of the mutants did. We then did an IPTG concentration gradient, growing the cells in liquid cultures, to see if there was an optimal concentration for expression, and if the mutants required more or less IPTG than the wild type for expression. Using total concentrations of 0.05 mM, 0.1 mM, 0.25 mM, 0.5 mM, and 1.0 mM, we found that the cells expressed the wild type purple in every concentrations (with relatively equal intensity), while none of the mutants turned purple at any of the concentrations.

Testing for Oligomerization States

In order to test for the relative sizes/oligomerization of the mutant and wild type proteins, we first had to grow large cultures of the cells that contained the appropriate plasmids and purify the proteins from them. We decided to do so with only the single mutants and the triple mutant to see what the effects of each mutation were, and if the combination of all mutations caused some change that they did not do individually. The overnight culture of mutation R149E, however, did not grow, so it is excluded from these results.

After purification of the protein through the Ni-NTA columns, we collected five 200 µl fractions containing the protein. Based on the control, we chose the two fractions that appeared to contain the most protein (e.g., the two fractions from the control that were the most intensely purple were the fractions we used for each mutant). These fractions were numbers two and three. We then ran each of the samples we had collected from the purification on protein gels, including a whole cell sample, the soluble protein sample after centrifugation, the flow through from the column, and fractions two and three for each protein. It is clear from the gel [Figure 6] that the wild type purple protein is at a much higher concentration than the mutants. We can also see that the mutants have dark bands at the same position that the purple protein showed up in the control, and the triple mutant is significantly lower in concentration than the others.

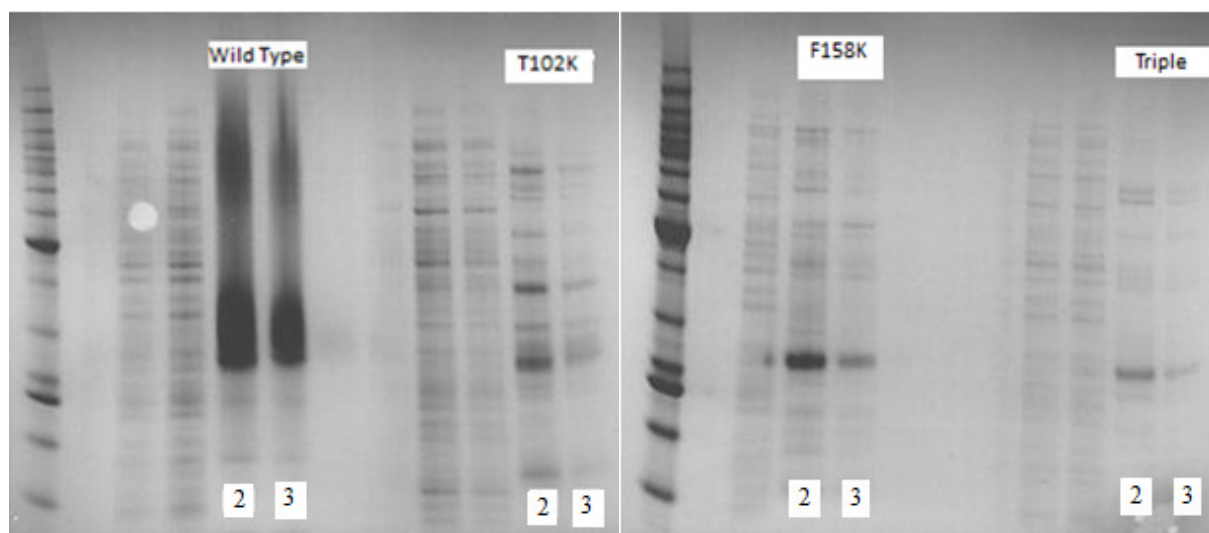


Figure 6. Protein gels of fractions 2 and 3 of each of the wild type purple protein, mutation T102K, mutation F158K, and the triple mutant. The top label for the lanes indicates the mutant (or wild type) that the samples come from, and the important lanes under each heading are the two labeled “2” and “3” for fractions 2 and 3, respectively.

Sephacryl S200 Column Chromatography

One fraction of each of the mutant and wild type proteins were passed through the Sephacryl S200 Column. The column separated the proteins in each sample by relative molecular weight, eluting the proteins into 60 2.5-ml fractions. As the samples passed through the column and were eluted into the fractions, the relative amounts of protein coming out at each point were recorded onto a chart. The data showed a large peak during the elution for the wild type protein, indicating the time when it eluted from of the column, and thus giving us a size standard for comparison [Figure 7].

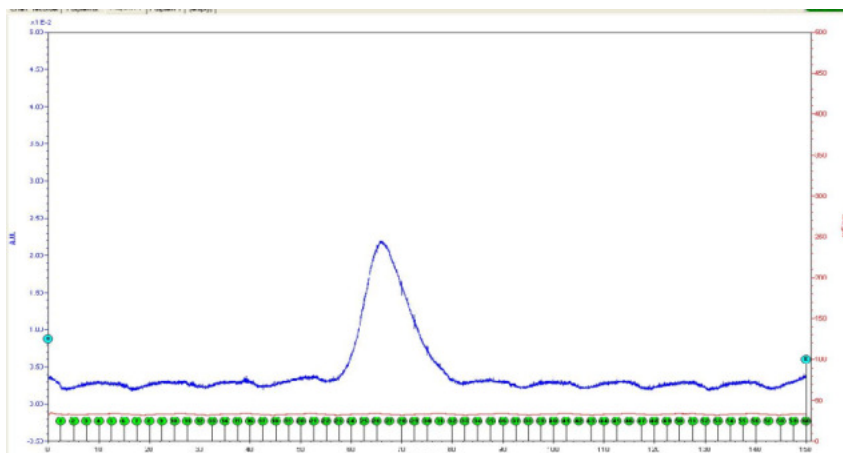


Figure 7. Sephadex S200 Column recording for wild type purple protein.

In the figure, the elution fractions are marked along the x-axis (of time, in minutes), and the absorbance reading is on the y-axis. While the control showed a clear and significant peak, the readings for all of the three mutants showed no significant peaks, and their charts were comparable to that of buffer being run through the column. Therefore, there was insufficient material to determine from the chart what the relative oligomerization states of the mutants were.

Western Dot Blot

Because the protein level was too low for the Sephadex S200 to give clear results, the western dot blot was done as a more sensitive alternative test. By using antibodies and chemiluminescence, we were able to view the relative concentrations of protein across the 60 fractions by looking at how intensely the dots for each sample are across the gradient. For the dot blot, we used every third fraction (so a total of 20 samples) as our interval. From the image [Figure 8], the gradient for the wild type purple protein seems to have two “peaks” of brightness. However, the second peak can be considered degraded protein because it is around the last fractions to be eluted and did not show up on the Sephadex chart. The area that is important – the first bright peak – appears around 50 ml during the elution (the center dot of the peak is

circled in red). Based on a size standard, we were able to calculate that, according to that data, the purple wild type protein is about 65 kDa in size.

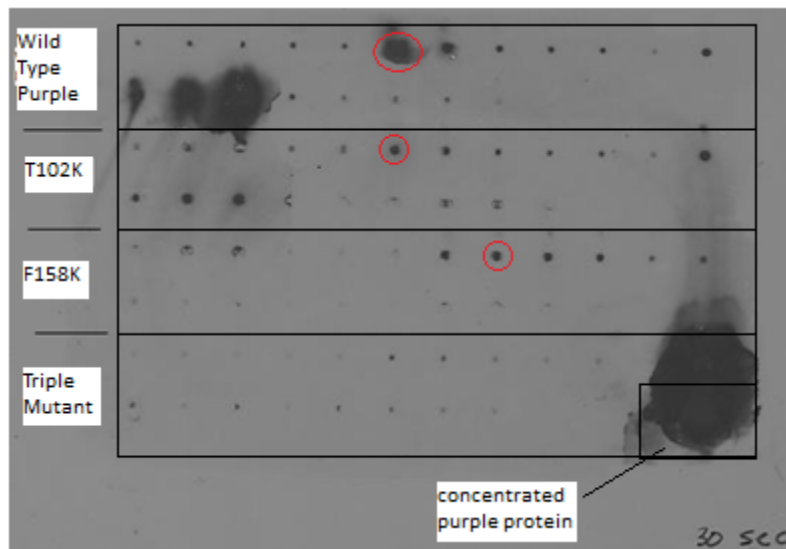


Figure 8. Western Dot Blot of wild type purple protein and mutant proteins. The dots represent every third fraction, starting from the top left and ending at the bottom right of each box. The red circles indicate the approximate centers of the peaks where the fractions gave the greatest signal. X-ray done using 30 second exposure.

Continuing to look at the X-ray image, we can see that mutation T102K has the same major peak as the wild type above it, indicating that it is relatively the same molecular weight. However, looking at mutation F158K, it appears that the peak has shifted downstream significantly (see that the red circle around the approximate center of the peak is not at the same location as the other mutation or the wild type purple). By comparing this to a standard, we can tell that the mutant containing F158K is about 32 kDa – therefore, it is about half the size of the purple wild type. For the triple mutant at the bottom of the image, the expression is too low to indicate a signal anywhere, so nothing can be concluded about its size. (Recall that on the protein gel in Figure 4, the triple mutant had the faintest band.)

Discussion

According to the results, both mutations T102K and F158K affected the spectral qualities of purple protein and prevented cells from expressing the purple color. Due to the low signal from the protein purification experiments, this may partially be due to the lower expression of purple protein itself. In fact, during the purification process, mutation T102K exhibited a pale purple color on the Ni-NTA columns, indicating a low level of purple, and this mutant was the one with the highest level of expression (after the wild type purple). It could also involve the mutations causing a change in the chromophore itself, such as its structure or maturation time.

With regard to oligomerization state, mutation T102K was unchanged during the experiment, as shown by the western dot blot test. Mutation F158K, on the other hand, altered the oligomerization state of the purple protein. From the dot blot, we found that the peak for this mutation's elution from the Sephadex S200 column showed it to be half the size of the wild type protein. This leads us to several conclusions, one being that the oligomerization state of purple protein is greater than one. We can also conclude that, using this specific mutation, we can reduce the oligomerization state, making it a more manageable size for protein tagging experiments. However, we cannot definitively say whether the change in oligomerization state was a shift from a tetramer (wild type) to a dimer (F158K mutant) or a shift from a dimer to a monomer. Further experimentation, for example using analytical ultracentrifugation, will allow us to definitively determine the mass and oligomerization state of these proteins.

In its current state, the reduced purple protein is still not as useful as a protein tagging tool as we would like. However, future research will be done to recover the spectral properties the F158K mutant. Through this project, we were able to create a protein of a more manageable structure that may eventually be a new color of protein fusion tag. From here, random

mutagenesis experiments should be done using the mutant as a template. Once they are transformed into high-efficiency protein expression cells, the purple colonies can be selected and screened for oligomerization state (through protein purification and filtration). If a successful clone can be found, then by purifying its DNA and attaching it to a protein of interest, it can be determined whether purple protein may be a useful addition to the rainbow of protein tags.

Appendix

Purple Wild Type and Mutant Sequences

* Locations of the mutations made are highlighted. See KEY below for which mutation each color indicates.

Purple Wild Type:

AGGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCGAAGGCGATGGA
AAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAAGCTCACTGTCACCAAGGGTG
GACCTCTGCCATTTGCTTGGGATATTTTATCACCCTGTCTCAATACGGAAGCATACC
ATTCACCAAGTACCCTGAAGACATCCCTGATTATGTAAAGCAGTCATTCCCTGAGGG
ATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGTACTGTCAGCA
ATGATTCCAGCATCCAAGGCAACTGTTTCATCTACAATGTCAAAATCTCTGGTGTGA
ACTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAAC
ACTGAGCGTCTCTTTGCA^{CGA}GATGGAATGCTGATAGGAAACAAC^{TTT}ATGGCTCTG
AAGTTGGAAGGAGGTGGTTACTATTTGTGTGAGTTCAAATCTACTTACAAGGCAAAG
AAGCCTGTGAGGATGCCAGGGTATCACTATGTTGACCGCAAACCTGGATGTAACCAG
TCACAACAAGGATTACACATTTGTTGAGCAGTGTGAAATATCCATTGCACGCCACTC
TTTGCTCGGTCATCACCATCACCATCACTAA

Mutations Made:

AGGTTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCGAAGGCGATGGA
AAAGGAAAGCCTTACGAGGGGGAGCAGACGGTAAAGCTCACTGTCACCAAGGGTG
GACCTCTGCCATTTGCTTGGGATATTTTATCACCCTGTCTCAATACGGAAGCATACC
ATTCACCAAGTACCCTGAAGACATCCCTGATTATGTAAAGCAGTCATTCCCTGAGGG
ATATACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTGT^{aaa}GTCAGCAA
TGATTCCAGCATCCAAGGCAACTGTTTCATCTACAATGTCAAAATCTCTGGTGTGAA
CTTTCCTCCCAATGGACCTGTTATGCAGAAGAAGACACAGGGCTGGGAACCCAACA
CTGAGCGTCTCTTTGCA^{gaa}GATGGAATGCTGATAGGAAACAAC^{aaa}ATGGCTCTGAAG
TTGGAAGGAGGTGGTTACTATTTGTGTGAGTTCAAATCTACTTACAAGGCAAAGAAG
CCTGTGAGGATGCCAGGGTATCACTATGTTGACCGCAAACCTGGATGTAACCAGTCAC
AACAAGGATTACACATTTGTTGAGCAGTGTGAAATATCCATTGCACGCCACTCTTTG
CTCGGTCATCACCATCACCATCACTAA

KEY: ^{Mutation T102K}
^{Mutation R149E}
^{Mutation F158K}

Primer Sequences

Mutation R149E

Forward: ACTGAGCGTCTCTTTGCA^{gaa}GATGGAATGCTGATAGGA

Reverse: TCCTATCAGCATTCCATCTTCTGCAAAGAGACGCTCAGT

Mutation F158K

Forward: ATGCTGATAGGAAACAAC^{aaa}ATGGCTCTGAAGTTGGAA

Reverse: TTCCAACCTCAGAGCCATTTTGTGTTTCCTATCAGCAT

Mutation T102K

Forward: GAAGATGGTGCAGTGTGT^{aaa}GTCAGCAATGATTCCAGC

Reverse: GCTGGAATCATTGCTGACTTTACACACTGCACCATCTTC

References

- Baird, G. S., Zacharias, D. A., & Tsien, R. Y. (2000). Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences of the United States of America* , 97 (22), 11984-9.
- BD Biosciences. (2005, Jan). *BD Living Colors DsRed-Monomer Fluorescent Protein*. Retrieved Mar 20, 2009, from Clontech:
http://wolfson.huji.ac.il/purification/PDF/Tag_Protein_Purification/FluorescentProteins/BD_DsRedMonFluorProt.pdf
- Giepmans, B. N., Adams, S. R., Ellisman, M. H., & Tsien, R. Y. (2006). The fluorescent toolbox for assessing protein location and function. *Science* , 312 (5771), 217-24.
- Gurskaya, N. G., Fradkov, A. F., Tersikh, A., Matz, M. V., Labas, Y. A., Martynov, V. I., et al. (2001). GFP-like chromoproteins as a source of far-red fluorescent proteins. *FEBS Letters* , 507 (1), 16-20.
- Labas, Y. A., Gurskaya, N. G., Yanushevich, Y. G., Fradkov, A. F., Lukyanov, K. A., Lukyanov, S. A., et al. (2002). Diversity and evolution of the green fluorescent protein family. *Proceedings of the National Academy of Sciences of the United States of America* , 99 (7), 4256-61.
- Lukyanov, K. A., Fradkov, A. F., Gurskaya, N. G., Matz, M. V., Labas, Y. A., Savitsky, A. P., et al. (2000). Natural Animal Coloration Can Be Determined by a Non-fluorescent Green Fluorescent Protein Homolog. *Journal of Biological Chemistry* , 275 (34), 25879-82.
- Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A. P., Zeraisky, A. G., Markelov, M. L., et al. (1999). Fluorescent proteins from Nonbioluminescent Anthozoa species. *Nature Biotechnology* , 17 (10), 969-73.
- Matz, M. V., Lukyanov, K. A., & Lukyanov, S. A. (2002). Family of the green fluorescent protein: journey to the end of the rainbow. *Bioessays* , 24 (10), 953-9.
- Pakhomov, A. A., & Martynov, V. I. (2008). GFP Family: Structural Insights into Spectral Tuning. *Chemistry and Biology* , 15 (8), 755-64.
- Schultz, C., & Kohn, M. (2008). Simultaneous Protein Tagging in Two Colors. *Chemistry and Biology* , 15 (2), 91-2.
- Shaner, N. G., Patterson, G. H., & Davidson, M. W. (2007). Advances in fluorescent protein technology. *Journal of Cell Science* , 120 (Pt 24), 4247-60.

- Tasdemir, A., Khan, F., Jowitt, T. A., Iuzzolino, L., Lohmer, S., Corazza, S., et al. (2008). Engineering of a monomeric fluorescent protein AsGFP499 and its applications in a dual translocation and transcription assay. *Protein Engineering, Design and Selection*, 21 (10), 613-22.
- Yang, F., Moss, L. G., & Phillips, G. N. (1996). The Molecular Structure of Green Fluorescent Protein. *Nature Biotechnology*, 14 (10), 1246-51.

Acknowledgments

I would first and foremost like to thank Dr. Scott Stevens for being an excellent research advisor. His help in the planning and execution of this project has been invaluable. I have learned a great deal from him and could not have asked for a better supervisor. He has also been a vital part of the actual writing of this thesis, and I truly appreciate all the time Dr. Stevens has devoted to making this project a success.

I also want to thank Matthew Sorenson for providing daily guidance and assistance during the course of our research. Matt, Jennifer Hennigan, and Rea Lardelli were always willing to take time out of their own work to help us when we needed it. Leodis (Champ) Gupton also made sure that we had the supplies we needed (such as media plates) for the various steps involved in the project.

Additionally, I would like to thank Dr. Albert MacKrell for preparing me with the necessary skills and knowledge over the past years to approach this project with understanding and ability. Finally, I want to thank my parents and friends for supporting me through the stress and time commitment involved in the research for this project.